(4)

UNCLASSIFIED

CENTRAL STUDIES BRANCH
Canberra A.C.T.

CSB MEMORANDUM 3

MAY 1988

AR-005-257

SOME STATISTICAL ASPECTS OF ATTRITION STUDIES

BY B.K. McMILLAN

APPROVED FOR PUBLIC RELEASE

AUSTRALIA

Department of Defence

COPY NO. 45 OF 48

UNCLASSIFIED

88

CSB MEMORANDUM 3

MAY 1988

SOME STATISTICAL ASPECTS OF ATTRITION STUDIES

B.K. McMILLAN

ABSTRACT

This memorandum gives a methodology for analysing attrition data, and establishing the reliability of the results. Underlying assumptions include exponentially distributed times between attritions and a mean that increases with operating experience. Principal techniques used are Maximum Likelihood Estimation and Jackknifing. An example is given, drawn from F/A-18 aircraft data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

1. This paper is an outcome of studies in Aircraft Attrition, the bulk of which were completed in early 1985. It deals with some of the theoretical aspects of using a limited amount of data to evaluate the parameters of a commonly accepted 'learning curve' model, and to determine the reliability of those values. The statistical techniques used are not 'state of the art', but equally are not rudimentary. Computer assistance in the application of these aspects can be considered highly desirable.

2. The work documents and demonstrates an approach to Attrition modelling using Maximum Likelihood Estimation (MLE) and Jackknifing with a worked example from attrition modelling for the F/A-18 aircraft. Use of MLE is highly recommended for situations in which Normal distributions are not applicable. Jackknifing is an additional technique for improving the reliability of parameter estimates by reducing both bias and the effect of outliers. In addition it can allow error bounds on the estimates of single parameters to be approximated.

Rationale

3. While this paper has been written principally for the attrition analyst, it is appropriate to outline the reasons for adopting this particular approach.

4. Present practice in some areas is to fit (by least squares regression) a straight line to cumulative attrition data. Unfortunately, classical line fitting processes require that each data point be substantially independent of all others. Cumulative data, by its very nature cannot be independent since a change in one value will alter all those values that come after it. To overcome this problem given cumulative data, the differences between adjacent observations may be used instead. In attrition terms, this is the time between losses.

5. Another practice in some areas is to use least squares for regression or point estimation to determine the mean or trend of time between losses. Unfortunately the least squares approach requires that the variation of the data about the point or trend line be generally symmetric and preferably approximately Normally distributed. Attrition data however tend to have exponentially distributed times between losses. This is a phenomenon that has been observed in practice, and which is supported theoretically under the proposition that the probability of loss in any one short time interval is the same as in any other, for periods when learning effects can be neglected. The exponential distribution as it stands or even when transformed by taking logs, is not symmetric.

6. The approach adopted for this paper has been to assume that attrition intervals are (locally) exponentially distributed but with a mean that increases with flying experience. While an exponential learning process has been used, the more general Weibull distribution has not been adopted since it restricts theoretical learning process options. With some additional work the analyst may follow the general lines of this paper using another form of learning such as exponential to an asymptote etc. A frequently used methodology for handling parameter estimation in non-Normal situations is that of Maximum Likelihood Estimation (MLE) and is adopted here.

7. Data from distributions which have long tails (e.g. the exponential) can sometimes have atypical values (outliers) that affect the estimate of the parameter required. The effect of outliers can be reduced using a technique called Jackknifing. In addition, Jackknifing produces intermediate values which are more likely to be approximately Normally distributed that the original data. When this is the case, it allows classical statistical methods to be used in estimating the accuracy with which the single parameters have been found. Other methods for determining accuracy are often difficult to apply. A further advantage of Jackknifing is that it reduces any bias that is intrinsic in the estimation process.

Assumptions for Attrition Modelling

8. Attrition studies generally assume that over short periods each individual attrition or 'event' is independent of any other and that events occur with equal probability in equal time intervals. From this it follows that the Poisson distribution describes the numbers of events occurring in set time intervals or, equivalently, that the times between successive events are Exponentially distributed (over short periods).

9. Another common assumption is that over long periods the attrition rate decreases with increasing operational experience, and that this decrease is exponential, as for example in the following equation:

$$E = Ah^{-B} \tag{1}$$

where E is the expected number of events per unit of operating time, (e.g. attritions per flying hour) and h is the cumulative operating time. A, B are constants requiring statistical evaluation. This can represent a form of learning where, with experience, fewer losses occur. The learning process is generally fairly slow, however, and may even depend on the attritions themselves (i.e.

Equation 1 could be an approximation of a series of steps). The relatively slow learning is assumed to be such that infringement of the Poisson/Exponential requirements can be neglected. This particular approach to the problem also has considerable flexibility in alternate model selection since different forms of learning equation can be adopted, using the same general approach (e.g. exponential learning to an asymptote).

10. Integrating Equation 1 from time h_o to h gives the expected <u>cumulative</u> number of events (C) that have occurred up to time h:

$$C = \frac{A\ (h^{1-B} - h_o^{1-B})}{1 - B} \tag{2}$$

Since it is a cumulative value, C must be positive and continuously increase, either to an asymptote (1-B negative) or indefinitely (1-B positive). In addition if the number of expected events per unit time is decreasing (i.e. 'learning' is faster than 'burnout') then B must be positive, although negative values for B do have a physical significance.

11. The other constant (A) in Equations 1 and 2 must also be positive since both cumulative operating time (h) and number of events must be positive.

12. In summary then, the assumptions are:

a. over short periods, times between events are well approximated by the exponential distribution and thus numbers of events per unit time approximate the Poisson distribution;

b. learning applies in an exponential manner (Equation 1); and is a factor in determining the mean of the exponential distribution at any given time; and

c. constant A of Equation 1 must be positive.

13. Although the Poisson distribution and the cumulative events model (Equation 2) are frequently encountered in attrition modelling, they are not really satisfactory as statistical models since:

a. in this context the Poisson distribution requires an artificially defined unit of time within which the number of events must be counted; and

b. in the cumulative model, sequential observations are not independent.

To avoid such problems, time between events may be used instead, and therefore the model to determine the essential parameters A and B is:

$$m = \frac{h^B}{A} \tag{3}$$

and $$p(t) = \frac{e^{-t/m}}{m} \tag{4}$$

where A and B are constants as before,

h is the cumulative operating experience,

m is the mean time between events,

t is the time between events, and

p(t) is the probability density function of t.

14. Observations of inter-event times are generally numbered in chronological order. This numbering, represented by 'i' is then used to index the model variables (m, t and h). Thus the time between events is related to cumulative time through:

$$t_i = h_i - h_{i-1} \tag{5}$$

where $h_0 = 0$ usually

and Equations 3 and 4 can be rewritten:

$$m_i = \frac{h_i^B}{A} \qquad (6)$$

$$p(t_i) = \frac{\exp(-t_i/m_i)}{m_i} \qquad (7)$$

where h_i is the cumulative operating experience at event i,

m_i is the mean interevent time at event i, and

t_i is the time between events i-1 and i.

15. Thus Equations 5 to 7 represent the statistical model of exponential learning (Equation 6) with an exponentially distributed time between events.

Initial Parameter Estimation by MLE

16. It is common practice to assess the parameters of models using the method of Least Squares. This method works well when the distributions underlying the data are Normal or even when they are just symmetrical. However it has been found that skew distributions such as the exponential do not respond so well. In this case an alternative methodology can be evolved using Maximum Likelihood Estimation (MLE) techniques. When applied to fitting straight lines through data distributed normally across the line, MLE gives the same result as the Least Squares method. As its name implies, MLE yields parameter value estimates which maximize the likelihood of occurrence of the set of observed values, given the underlying distribution.

17. Following the techniques for MLE outlined in Reference 4, the method for this problem is to obtain analytically the overall probability of all the n observations occurring. That is to maximize

$$P = p_1 p_2 \cdots p_i \cdots p_n$$

where p_i are distributed as $p(t_i)$ in Equation 7 and P is to be maximized with respect to A and B. It can be shown that this is equivalent to maximizing:

$$Q = \ln(P) = \sum_i \ln(p_i)$$

$$= \sum_i (-t_i/m_i - \ln(m_i)) \tag{8}$$

which can be solved for A and B (implicit in m_i) by using two simultaneous equations obtained from partial differentiation. For a maximum, the partial derivatives must be zero:

$$\frac{\partial Q}{\partial A} = 0$$

$$\frac{\partial Q}{\partial B} = 0 \tag{9}$$

It should be noted that these derivatives can be expanded, so:

$$\frac{\partial Q}{\partial A} = \sum_i \left(\frac{\partial Q}{\partial m_i} \frac{\partial m_i}{\partial A}\right) = 0 \tag{10}$$

and $$\frac{\partial Q}{\partial B} = \sum_i \left(\frac{\partial Q}{\partial m_i} \frac{\partial m_i}{\partial B}\right) = 0$$

Evaluation of the derivatives can therefore proceed by partial differentiation of Q with respect to m_i, and m_i with respect to A and B. Differentiation of Equation 8 with respect to m_i yields:

$$\frac{\partial Q}{\partial m_i} = \sum_i \left(\frac{-1}{m_i} + \frac{t_i}{m_i^2}\right) \tag{11}$$

and differentiation of Equation 6 with respect to A and B yields:

$$\frac{\partial m_i}{\partial A} = \frac{-h_i^B}{A^2} = \frac{-m_i}{A}$$

$$\frac{\partial m_i}{\partial B} = h_i^B \frac{\ln(h_i)}{A} = m_i \ln(h_i) \tag{12}$$

Thus the two simultaneous Equation 10 are:

$$\sum_i \left(\frac{-1}{m_i} + \frac{t_i}{(m_i)^2}\right)\frac{m_i}{A} = 0$$

$$\sum_i \left(\frac{-1}{m_i} + \frac{t_i}{(m_i)^2}\right) m_i \ln(h_i) = 0 \tag{13}$$

which may be simplified to:

$$n = \sum_i \frac{t_i}{m_i}$$

and $$\sum_i \ln(h_i) = \sum_i \left(\frac{\ln(h_i) t_i}{m_i}\right) \tag{14}$$

Expanding m_i as in Equation 6 allows these equations to be rewritten as:

$$A = \sum_i \frac{n}{t_i h_i^B} \tag{15}$$

$$A = \frac{\sum_i \ln(h_i)}{\sum_i \left(\frac{t_i \ln(h_i)}{h_i^B}\right)} \tag{16}$$

Evaluation of A and B proceeds by selection of a value for B, which is used to evaluate A through each of Equations 15 and 16. The difference between these two A values is then used to predict a new B, aiming for zero difference in the A values. The process is then repeated until Equations 15 and 16 give the same results.

18. In the event that B is defined by some other method or process, (for example through data from aircraft of the same generic type) Equations 9 to 14 still apply, thus allowing Equation 15 to be used to evaluate A directly.

<u>Improving and Assessing Parameter Estimates by Jackknifing</u>

19. The preceding section has shown how to estimate the parameters A and B, given a set of data. The process of Jackknifing, in general terms, re-assesses such parameters, using only a reduced data set, but for all possible reduced data sets. The name of the technique is said to have arisen from the fact that the blade of a Jackknife folds back into itself, just as the set of original observations are folded back into themselves. For example, a data set, of say n observed events, reduced by one observation, will give n reduced subsets of size n-1 observations. These subset estimates are then suitably combined with the original estimate (from the whole data set) to give an improved parameter estimate. Also of importance however, is the fact that the distribution of the subset estimates is often such that Student's t distribution can be applied to give a first approximation to the confidence limits on the improved estimate. It should be noted that there is no theoretical assurance of the applicability of Student's t in all circumstances, but reasonable results appear likely in many practical situations. Maximum likehood estimation is thought to be particularly suitable for use with Jackknifing. The approach used here is based on that of Reference 1.

20. A few points need to be mentioned about Jackknifing and extreme points. The gross impact of the Jackknife in reducing the effect of occasional extreme points on the estimate should not be underrated. For example, given 20 data points the chance that all of them will lie inside the 95 per cent band is only about 1/3, (i.e. 0.95^{20}), and those outside it may abnormally bias the simple estimate. Note, however, that estimates which utilize points from the extremes for their result, or which take no values from such extremes (e.g. order statistics), may not respond well to Jackknifing. Conversely, estimates such as means, variances etc., for skew distributions with long tails should to respond better to the technique. This latter case is the region into which attrition modelling fits.

21. Briefly, the definitions and equations required to Jackknife by dropping single observations are detailed below. A simplified and modified form of the notation of Reference 1 is used in order to clarify the presentation:

a. A parameter Θ is required (e.g. mean, standard deviation, constant in a regression line, etc).

b. There is a data set of observations, each complete observation identified by $\underline{x}_i$ ($i = 1$ to n) and the whole set identified by $\underline{X}$. Subsets are identified by $\underline{X}_{-i}$ which represents $\underline{X}$ less observation $\underline{x}_i$.

c. t_o is the estimator of Θ from $\underline{X}$.

d. t_{-i} the estimator of Θ from $\underline{X}_{-i}$.

e. The Jackknifed estimate of Θ is t^*:

$$t^* = nt_0 - \frac{n-1}{n}\sum_i t_{-i} \tag{17}$$

If confidence intervals are required then additional definitions are needed:

f. So called 'pseudo-values':

$$t^*_{-i} = nt_0 - (n - 1)t_{-i} \tag{18}$$

which relate to the Jackknifed estimate of Θ as

$$t^* = \frac{\sum_i t^*_{-i}}{n}$$

g. The variance of t^* is $var(t^*)$:

$$var(t^*) = \frac{\left(\sum_i (t^*_{-i})^2 - n(t^*)^2\right)}{(n - 1)}$$

which is the conventional definition of a sample variance for observations t^*_{-i}. This may be simplified to:

$$var(t^*) = n(n - 1)\left(\frac{\sum_{i=1}^{n} t^2_{-i}}{n} - \left(\frac{\sum t_{-i}}{n}\right)^2\right) \tag{20}$$

h. Application of the Student's t distribution is subject to the distribution of t_{-i} being approximately Normal, using n-1 degrees of freedom for the n values of t_{-i} and is to

$$\frac{(t^* - \Theta)\sqrt{n}}{\sqrt{var(t^*)}} \tag{21}$$

In practice the pseudo values are simply scaled and translated t_{-i} (Equation 18) so Normality in the t_{-i} is all that needs to be approximated.

22. Should it be necessary to estimate confidence intervals on more than one parameter, likelihood ratio tests may be considered. If it is felt that the pseudo-values are drawn from a multivariate Normal distribution, the likelihood ratio tests lead to the T^2 test as an analogue of the t-test. Details may be found in Reference 6.

23. There are some other very broad conditions which should be met for the technique to be applicable. Other than a differentiability condition, and to say that most continuous parametric statistical situations meet the conditions, their discussion is beyond the scope of this paper (for more details see References 1 and 2). The differentiability condition is that t estimates a function that after differentiation is continuous. Amongst other things, this implies that the estimator be unconstrained, and therefore the constant A (from the learning model) cannot be Jackknifed. Constant B, while conceptually constrained can in fact be estimated at a value outside its conceptual constraints through an abnormal (but possible) data set. To solve the problem for A, the log of A can be used, giving multiplicative rather than additive confidence intervals. Solutions for the B problem will be specific to the situation, the data set and the certainty on the bounds of B. In many cases it may be acceptable that 'unlearning' can occur, especially in the case of ageing equipment, new operating practices or conditions, and new operators. The concept of an upper bound on the total number of events may also be acceptable, particularly if only a limited range of cumulative operating experience is to be modelled. Both logarithms and fractional (e.g. $x/(a + bx)$) transforms can be considered if necessary, but care should be taken as they will affect the confidence limits considerably.

Example

24. Attrition data for the F/A-18 on a world wide basis, excluding losses that are not flying related has been used. As much of this work was done in early 1985 when only five losses had been sustained, the data have been partioned at that point. Table 1 gives data for the first five losses and Table 2 for the succeeding 22 losses. The first set of data was used to predict the loss rate for the midpoint of the period covering the last 10 losses, and compared with the average loss rate for those losses. Parameters assessed are A and B in the equation

$$m = \frac{h^B}{A}$$

where m is the mean interevent time, and h is the cumulative flying experience (hours) (Equation 3). Studies of similar US aircraft indicate that a good a priori value for B is 0.247. Comparisons of results using Least Squares, MLE and Jackknifed MLE have been given, as well as the effect of the a priori information.

Table 1. FIRST FIVE F-18 LOSSES

LOSS (i)	CUMULATIVE OPERATING EXPERIENCE (h_i)	INTER EVENT TIME ($t_i = h_i - h_{i-1}$)
0	0	-
1	3337	3337
2	4379	1042
3	27300	22921
4	45805	18505
5	64717	18912

Table 2. F-18 LOSSES SIX ONWARD

LOSS (i)	CUMULATIVE OPERATING EXPERIENCE (h_i)	INTER EVENT TIME ($t_i = h_i - h_{i-1}$)	LOSS (i)	CUMULATIVE OPERATING EXPERIENCE (h_i)	INTER EVENT TIME ($t_i = h_i - h_{i-1}$)
6	75430	10713	17	389676	10133
7	124248	48818	18	397609	7933
8	143495	19247	19	404867	7258
9	154720	11225	20	456060	51193
10	198658	43938	21	462696	6636
11	248657	49999	22	472105	9409
12	304747	56090	23	490221	18116
13	347378	42631	24	490978	757
14	349723	2345	25	498631	7653
15	379543	29820	26	519600	20969
16	379543	0	27	519600	0

Equation 6 was suitably modified (by taking logs of A and h_i and estimating log m_i by log t_i) then Least Squares regression was applied, yielding:

$A = 0.734$ $B = 0.891$

Application of MLE yielded:

$A = 0.356 \qquad B = 0.831$

And using the Jackknife with the MLE estimates gave:

$A = 0.0571 \qquad B = 0.655$

Should a value of B be selected for a priori reasons then only one parameter needs to be estimated, allowing an estimate of the confidence interval for that parameter when Jackknifed. In this instance previous studies of United States Air Force (USAF) aircraft have shown that over hundreds of attritions for both single and twin engine F/A type aircraft the best estimate for B is 0.247. Using this value gives the following results:

A by least squares = 0.00142

A by MLE = 0.00101

A by MLE Jackknifed = 0.00096

and the 90 per cent interval (with 4° of freedom, $t \approx 2.1$) is: $0.0047 > A > 0.00036$

15. Repeating this work for the whole data set (Tables 1 and 2) produced the results shown in Table 3.

Table 3. RESULTS FOR ALL F-18 LOSSES

METHOD	A	B	A FOR B = 0.247
Least Squares	0.000015	0.203	0.00373
MLE	0.000538	0.191	0.00105
Jackknifed MLE	0.00108	0.249	0.00104
90% Confidence Interval, 26° freedom, t = 1.7	-	-	0.00077 (Lower) 0.00141 (Upper)

26. Comparing these figures with the earlier values based on Table 1 it can be seen that the least variation in results occurs in the Jackknifed MLE results, except for the one parameter case where the differences in the MLE case and the Jackknifed MLE are too small to be significant. Further, if the a priori value for B really is the 'true' value for B, then the Jackknifed results were closer to that value with both data sets.

27. Figure 1 shows the attrition data (Tables 1 and 2) as times between losses, and the two parameter lines generated by the three different methods using the Table 1 data. Although none of the lines produces an excellent result for the Table 2 data, it is not difficult to see from this that the Jackknifed line is the best predictor. Using the a priori information with Table 1 data improves the prediction as shown in Figure 2. It is difficult to see which line is best, based on the Table 1 data alone, but when all data are included with these lines, it is clear that the Jackknifed line is best for the following reasons:

a. the MLE and Jackknifed lines from Table 3, B = 0.247 (based on 27 losses) would be indistingu able from line J Figure 2 (based on five losses) . drawn on Figure 2; and

b. bearing in mind the shape of the exponential distribution and the fact that its median is closer to zero than its mean, then the Least Squares line is probably the poorer predictor because it has 13 points above the line and 14 below.

It should be noted that although the differences between the lines look small, the figures are drawn on log paper. Mean times between events as estimated by lines M and J are nearly half as much again as those estimated by line L in Figure 2. Table 3 indicates almost a fourfold decrease in the mean times estimate for the Least Squares result over the other two methods. These are factors which can have a major impact when considering attrition buys. A further advantage from using the Jackknifing method is the availability of confidence intervals on the estimates which have been illustrated in this example, by the 90 per cent confidence interval (Figure 2).

28. The cumulative losses prediction is also of interest and is plotted for the one parameter Jackknifed MLE method, in Figure 3 along with the data. It should be noted that the confidence interval relates to the line and not the data, and that a particularly 'good' or 'bad' run of losses early in the loss history will shift all the following losses in relation to the line and intervals - i.e. the data points are not independent in this form.

29. If confidence intervals are important it may be necessary to check that a Normal approximation to the psuedo values is appropriate. In practice it may be found that for attrition modelling the interval overestimates a little on one side and underestimates a little on the other. This may be due to the long tail of the exponential distribution. Since a conventional variance definition is used with t^*_{-i} it would seem reasonable to use non-parametric statistics as an alternative for interval estimation. Reference 5 gives some such methods.

30. A quick check on the appropriateness of an assumption of Normality can be done by sorting the data (t_{-i}) and plotting the result at equal intervals on Normal Probability paper. Figure 4 illustrates this for Table 1. An adequate straight line fit to the data indicates Normality. The values for all t_{-i} calculated for Table 1 data are shown in Table 4. Included in the table are columns j which give the sorted order for the preceding column.

Table 4. INTERMEDIATE JACKKNIFE RESULTS

INDEX (i)	ln(A) RESULTS		B RESULTS		ln(A) RESULTS FOR B = 0.247		COMMENTS
	t_{-i}	j	t_{-i}	j	t_{-i}	j	
0	-1.03	-	0.83	-	-6.90	-	Pure MLE Result - i.e. t_{-0} estimates ln(A) and B
1	2.66	5	1.18	5	-7.02	2	
2	-3.05	1	0.64	1	-7.09	1	
3	-1.52	2	0.76	2	-6.66	5	
4	-0.81	3	0.86	3	-6.82	4	
5	-0.15	4	0.93	4	-6.84	3	

Note.

1. Columns j give the sorted t_{-i} order.

CONCLUSION AND SUMMARY

31. This paper has briefly indicated a methodology that may be followed in analysing attrition data, that is preferable to classical least squares type approaches. It has also given the rationale and underlying assumptions behind the work and demonstrated (through an example on F/A-18 aircraft attritions) the sorts of differences in results that might be expected.

32. The use of MLE methods are strongly recommended in this area, and the use of the Jackknifing is also commended whether or not MLE is used. A program has been written in PASCAL for attritions (using MLE and Jackknifing) and is available through Reference 3.

REFERENCES

1. Eyland, A., 'Jackknife Techniques', The Statistical Society of Australia and the Australian Society for Operations Research, Sydney Chapter, A Symposium on Developments of Simulation Theory, pp. 45-55, November 1977.

2. Efron, Bradley, 'The Jackknife, the Bootstrap and Other Resampling Plans'.

3. McMillan, B.K., 'Program ASJMLEAT', Central Studies Branch, Department of Defence, Canberra, 1985.

4. Freund, John E., 'Mathematical Statistics', 1965.

5. Hollander and Wolfe, 'Nonparametric Statistical Methods'.

6. Anderson, T.W., 'An Introduction to Multivariate Statistical Analysis', John Wiley and Sons, 1966.

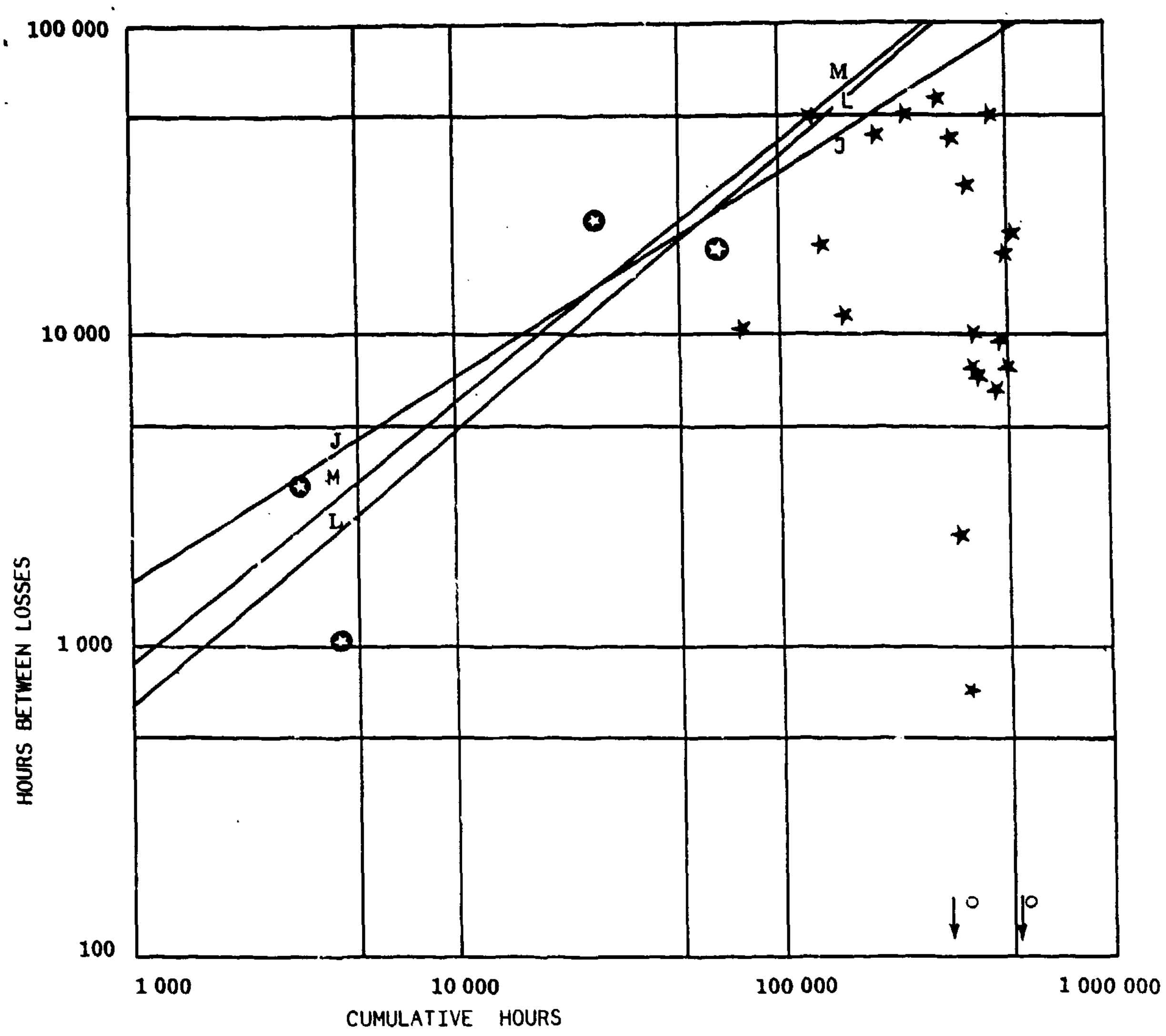

Figure 1. INTER-EVENT TIMES AND TWO PARAMETER LINES (FIRST FIVE LOSSES)

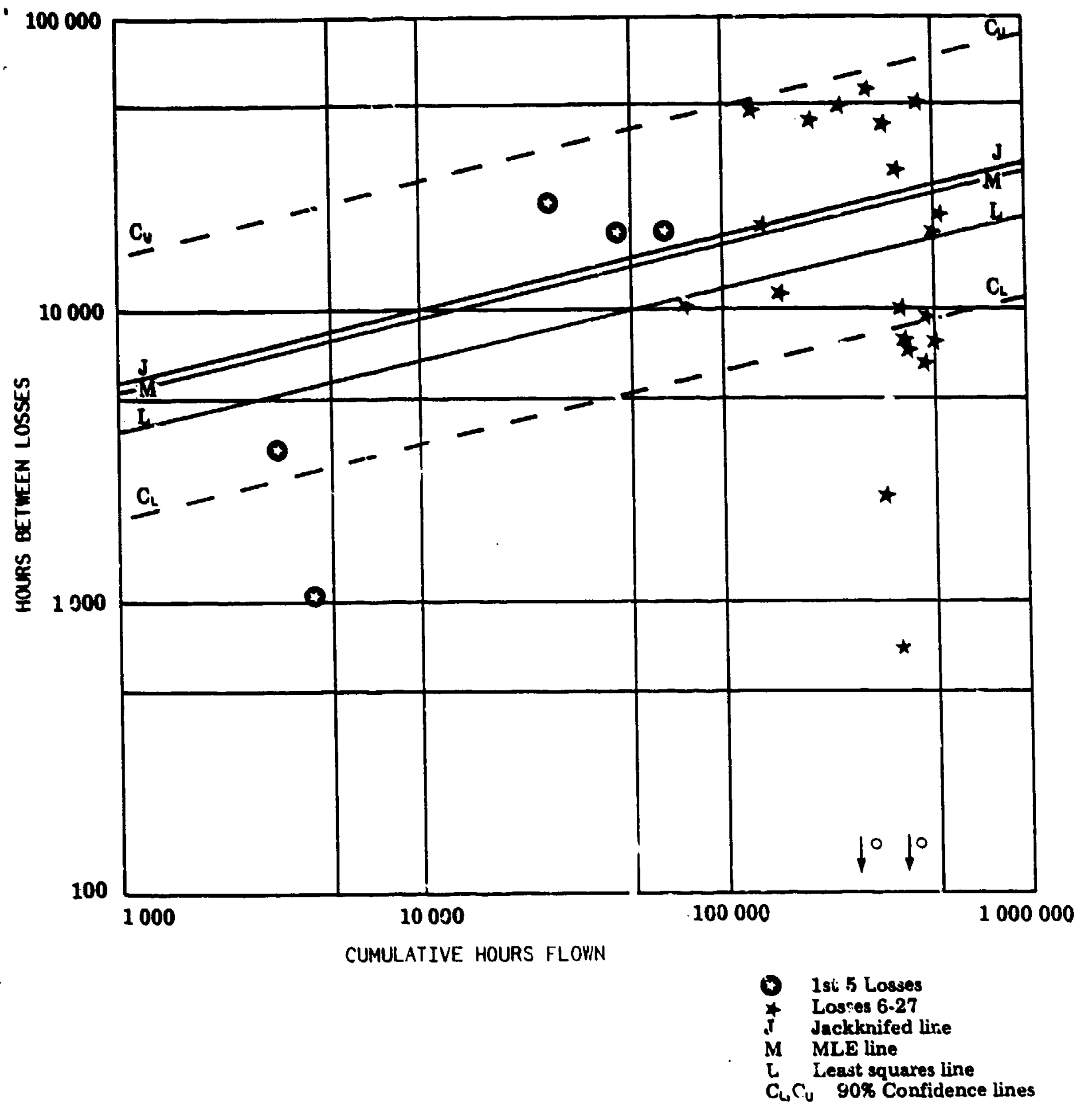

Figure 2. INTER-EVENT TIMES AND LINES FOR B = 0.247

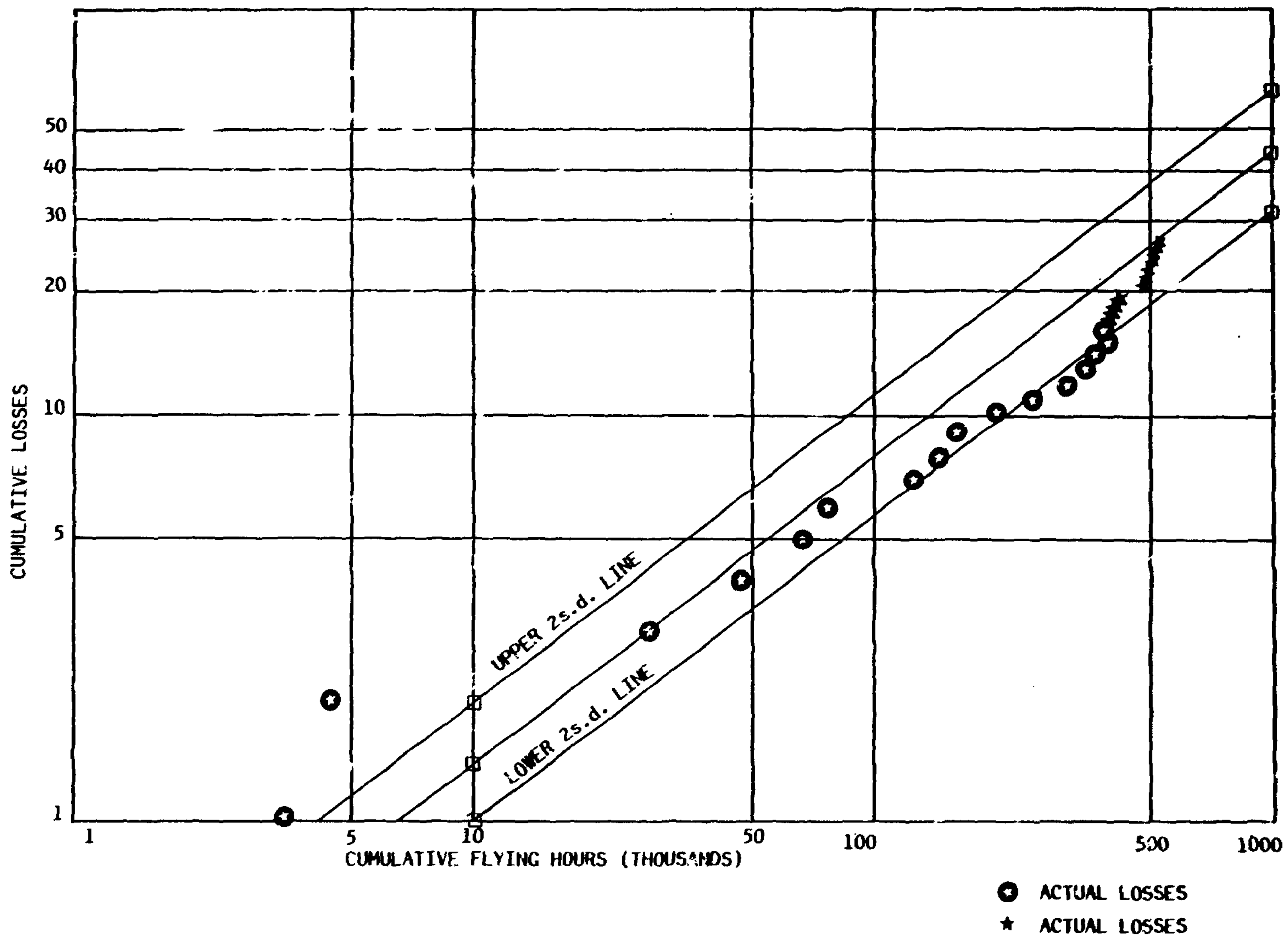

Figure 3. F/A-18 CUMULATIVE ATTRITION RESULTS

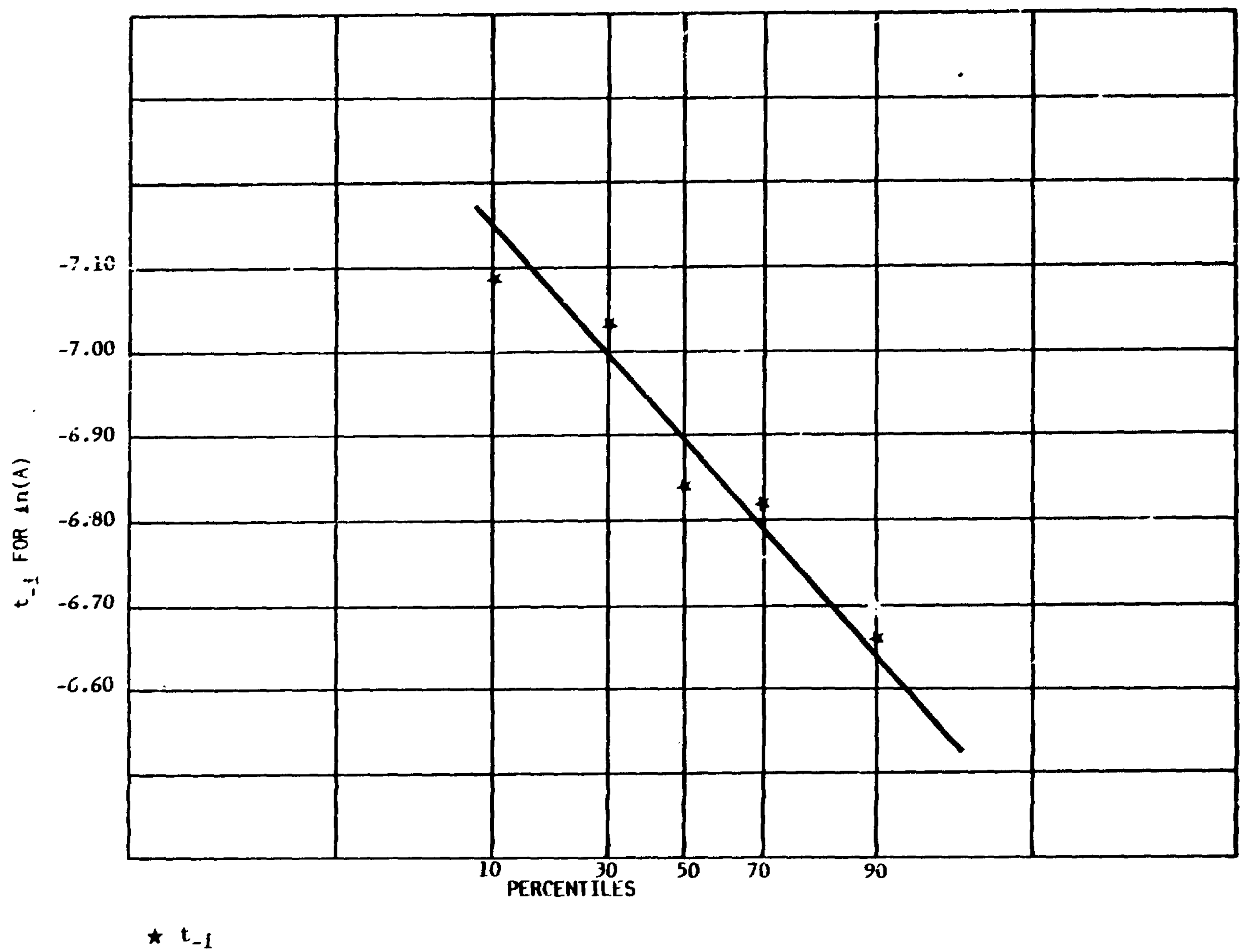

Figure 4. NORMALITY ASSESSMENT FOR t_{-i}, WHEN B = 0.247

DISTRIBUTION

	Copy No.
DEPARTMENT OF DEFENCE	
Central Office	
Deputy Secretary B	1
First Assistant Secretary, Force Development and Analysis Division	2
Headquarters Australian Defence Force	
Vice Chief of the Defence Force	3
Assistant Chief of the Defence Force (Policy)	4
Navy Office	
Navy Scientific Adviser	5
Director Operational Analysis - Navy	
Army Office	
Scientific Adviser - Army	6
Director Operational Analysis - Army	
Air Force Office	
Air Force Scientific Adviser	7
Director Operational Analysis - Air Force	
Defence Science and Technology Organisation	
Chief Defence Scientist	8
Assistant Chief Defence Scientist - Policy	9
Central Studies Branch	
Assistant Secretary, Central Studies	10
Principal Research Scientists, CSB (3)	11-13
Senior Service Representatives, CSB (3)	14-16
Author	17-26
Headquarters Land Command	
Staff Officer Science (Dr D. Conley)	27
Operational Research and Analysis Establishment, Canada	
Director General, Operational Research	28
Technical Report Centre (2)	29-30
Document Exchange Centre (18)	31-48

AL 149
Revised May 86

Department of Defence

DOCUMENT CONTROL DATA

Page Classification
UNCLASSIFIED

1a. A R Number	1b. Establishment Number	2. Document Date	3. Task Number
AR-005-257	CSB Memo 3	May 1988	

4. Title

SOME STATISTICAL ASPECTS OF ATTRITION STUDIES

5. Security Classification
(Place appropriate classification in box/s ie Secret(S), Confidential(C), Restricted (R), Unclassified (U))

U document U title U abstract

6. No. Pages

7. No. Refs

8. Author(s)

9. Downgrading/Delimiting Instructions

N/A

10. Corporate Author and Address

Central Studies Branch,
Department of Defence
CANBERRA, A.C.T. 2600.
AUSTRALIA

11. Office/Position responsible for

Sponsor CSB
Security CSB
Downgrading CSB
Approval CSB

12. Secondary Distribution *(Of this document)*

APPROVED FOR PUBLIC RELEASE.

Overseas enquiries outside stated limitations should be referred through ASDIS, Defence Information Services Branch, Department of Defence, Campbell Park, CANBERRA, ACT 2601

13a. This document may be announced in catalogues and awareness services available to..........

NO LIMITATIONS.

13b. Citation for other purposes *(ie. casual announcement)* may be ☐ unrestricted or ☐ as for 13a

14. Descriptors

Maximum Likelihood Estimation
Attrition
Jackknifing
Statistics
F/A-18 Data

15. COSATI Group

0072F

16. Abstract

A methodology for analysing attrition data and establishing the reliability of the results is given. Underlying assumptions include exponentially distributed times between attritions, and a mean that increases with operating experience. Principal techniques used are Maximum Likelihood Estimation and Jackknifing. An example is given drawn from F/A-18 data.